

Non è lecito utilizzare le registrazioni delle lezioni se non per motivi di studio individuale.
Recordings of online classes must be used for individual study purposes only.

[HOME](#) | [CORSI](#) | [STUDENTI LAUREE E LAUREE MAGISTRALI](#) | [A.A. 2021 - 2022](#) | [DIP. DI SCIENZE AMBIENTALI, INFORMATICA E STATISTICA](#)
| [LAUREE](#) | [CT3 - INFORMATICA](#) | [CT0429 \(CT3\) - 21-22](#) | [ESERCIZI](#) | [VECCHI APPELLI - ESERCIZI SUI GLM](#)

Iniziato	martedì, 15 novembre 2022, 11:23
Stato	Completato
Terminato	martedì, 15 novembre 2022, 11:24
Tempo impiegato	9 secondi
Valutazione	0,00 su un massimo di 21,00 (0%)

Domanda 1

Risposta non data

Punteggio max.: 4,00

Il responsabile della sicurezza di un sito vuole investigare l'ipotesi che l'esposizione mediatica dell'azienda ha un impatto sul numero di attività malevole verso il sito. Nell'arco di alcune settimane raccoglie quindi per un totale di 27 giorni le seguenti informazioni:

- Un indice che misura l'esposizione mediatica dell'azienda (variabile **exposure**)
- Il numero di attacchi informatici o in generale attività sospette (variabile **Nattacks**)

Inizia quindi la sua analisi stimando un modello lineare generalizzato (GLM) in cui la variabile **Nattacks** viene usata come variabile risposta e la variabile **exposure** viene usata come predittore:

```
fit1 <- glm(Nattacks ~ exposure, family = poisson)
```

```
summary(fit1)
```

```
Call:
glm(formula = Nattacks ~ exposure, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.34   -0.51   -0.09    0.39    1.58

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.457     0.150    9.7   <2e-16 ***
exposure       0.270     0.026   10.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 146.766  on 26  degrees of freedom
Residual deviance:  15.994  on 25  degrees of freedom
AIC: 140.9

Number of Fisher Scoring iterations: 4
```

Il modello viene usato per stimare il numero di attacchi in un giorno con un'esposizione mediatica pari a 5:

```
fit1$coefficients
nd1
```

```
(Intercept)    exposure
    1.4570      0.2699
```

```
exposure
1         5
```

```
predict(fit1, newdata = nd1, type="response")
```

```
[1] NA
```

Il responsabile desidera infine capire se il modello stimato è significativo rispetto ad un modello che non contiene nessun predittore.

Per tutte le analisi viene usato un livello di significatività $\alpha = 2\%$.

Si indichi quale dei possibili modelli corrisponde al modello effettivamente usato dall'analista. Si indichi come si può interpretare la stima ottenuta dal modello. Si indichi inoltre l'output mancante della funzione **predict**. Infine si indichi se il modello stimato dall'analista è significativo (rispetto al modello nullo).

- a. ☐ $Nattacks_i \sim Pois(\lambda_i)$.
☐ $Nattacks_i \sim Binom(n, p_i)$.
☐ $Nattacks_i \sim N(\mu_i, \sigma)$.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: $Nattacks_i \sim Pois(\lambda_i)$.

- b. ☐ Al crescere dell'esposizione cresce il numero degli attacchi.
☐ Al crescere dell'esposizione diminuisce il numero degli attacchi.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: Al crescere dell'esposizione cresce il numero degli attacchi.

- c. Si indichi il valore mancante dell'output della funzione `predict`.

✖

- d. ☐ Il modello è significativo.
☐ Il modello non è significativo.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: Il modello è significativo.

Il modello specificato è un modello Poisson che usa la funzione legame (link function) canonica, quindi la risposta corretta è:

$$Nattacks_i \sim Pois(\lambda_i), \quad \text{with } \lambda_i = \exp\{\beta_0 + \beta_1 \text{exposure}_i\}$$

La stima del coefficiente relativo ad `exposure` è positivo quindi al crescere dell'esposizione aumenta il numero degli attacchi.

La stima del numero di attacchi si ottiene con: $\exp\{1.457 + 0.2699 * 5\} = 16.55$.

Per verificare la significatività del modello si possono confrontare le devianze nulle e residue nel summary: un confronto formale richiede l'uso di un LRT (`1-pchisq(146.766-15.994, df=1)` = 0.00), ma la differenza tra le due devianze è così marcata che non vi è dubbio che il p-value del test sia molto piccolo.

- a. Vero / Falso / Falso
b. Vero / Falso
c. 16.55
d. Vero / Falso

Domanda 2

Risposta non data

Punteggio max.: 4,00

Un'agenzia immobiliare desidera indagare quali sono le caratteristiche che attirano maggior interesse nei clienti sul loro sito. Per un campione di 48 immobili vengono raccolte diverse caratteristiche e viene registrata l'informazione di quanti clienti contattano l'agenzia entro 24 ore dall'aggiunta dell'immobile sul sito. In particolare vengono registrate le seguenti informazioni:

- **n_cont**: il numero di clienti che hanno contattato l'agenzia entro 24 ore dall'aggiunta dell'immobile sul sito
- **n_foto**: il numero di foto inserite nel sito
- **m2**: la dimensione dell'immobile (in m²)
- **type**: l'informazione se l'immobile è una casa singola, una casa bifamiliare o un appartamento

La prima analisi si sofferma sull'effetto che ha il numero delle foto nell'annuncio (**n_foto**) sul numero attesi di contatti con l'agenzia:

```
dim(df)
```

```
[1] 48  4
```

```
fit_foto <- glm(n_cont~n_foto,family = poisson, data = df)
logLik(fit_foto)
```

```
'log Lik.' -80.899 (df=2)
```

```
deviance(fit_foto)
```

```
[1] 71.45
```

Si desidera usare il modello **fit_foto** per derivare una stima del predittore lineare del modello e del valore atteso di contatti in agenzia per un immobile con 30 foto:

```
coef(fit_foto)
```

```
(Intercept)      n_foto
      -2.578         0.203
```

```
predL <- predict(fit_foto, newdata = data.frame(n_foto = 30), type = "l")
predE <- predict(fit_foto, newdata = data.frame(n_foto = 30), type = "r")
```

```
predL
```

```
[1] NA
```

```
predE
```

```
[1] NA
```

Infine si stima un modello in cui tutte le variabili a disposizione sono inserite nel modello:

```
fit_all <- glm(n_cont~.,family = poisson, data = df)
```

```
deviance(fit_foto)
```

```
[1] NA
```

Si indichi il valore del BIC per il modello **fit_foto**. Si indichino poi le stime del predittore lineare del modello e del valore atteso di contatti in agenzia per un immobile con 30 foto. Infine si indichi se la devianza per il modello **fit_all** è maggiore, minore o uguale alla devianza per il modello **fit_foto**.

- a. Il valore del criterio BIC per il modello **fit_foto**:

✖

b. La stima del valore del predittore lineare predL: .

✖

c. La stima del valore del valore atteso predE: .

✖

d. ☐ deviance(fit_foto) = deviance(fit_all).☐ deviance(fit_foto) < deviance(fit_all).☐ deviance(fit_foto) > deviance(fit_all).

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: deviance(fit_foto) > deviance(fit_all).

BIC:

$$-2 * \log Lik(\mathcal{M}(p)) + \log(n) * p$$

Stima per il predittore lineare:

$$-2.5784 + 0.203 * 30$$

Stima per il numero atteso di contatti:

$$\exp(-2.5784 + 0.203 * 30)$$

La devianza funge la stessa funzione del RSS nei modelli lineari: più variabili si inseriscono nel modello più diminuisce la devianza.

a. 169.54

b. 3.51

c. 33.50

d. Falso / Falso / Vero

Domanda 3

Risposta non data

Punteggio max.: 5,00

Un'agenzia immobiliare desidera indagare quali sono le caratteristiche che attirano maggior interesse nei clienti sul loro sito. Per un campione di 34 immobili vengono raccolte diverse caratteristiche e viene registrata l'informazione se un possibile acquirente prende contatti con l'agenzia entro 24 ore dall'aggiunta dell'immobile sul sito. In particolare vengono registrate le seguenti informazioni:

- **cont**: variabile dicotomica con valore 0/1. Il valore 1 indica che un possibile acquirente ha preso contatti con l'agenzia
- **n_foto**: il numero di foto inserite nel sito
- **m2**: la dimensione dell'immobile (in m²)
- **type**: l'informazione se l'immobile è una casa singola, una casa bifamiliare o un appartamento

La prima analisi si sofferma sull'effetto che ha il numero delle foto nell'annuncio (**n_foto**) sulla probabilità che un possibile acquirente abbia preso contatti con l'agenzia:

```
fit_foto <- glm(cont ~ n_foto, data = df,
               family = binomial)
summary(fit_foto)
```

```
Call:
glm(formula = cont ~ n_foto, family = binomial, data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.853   -0.642   -0.280    0.872    1.610

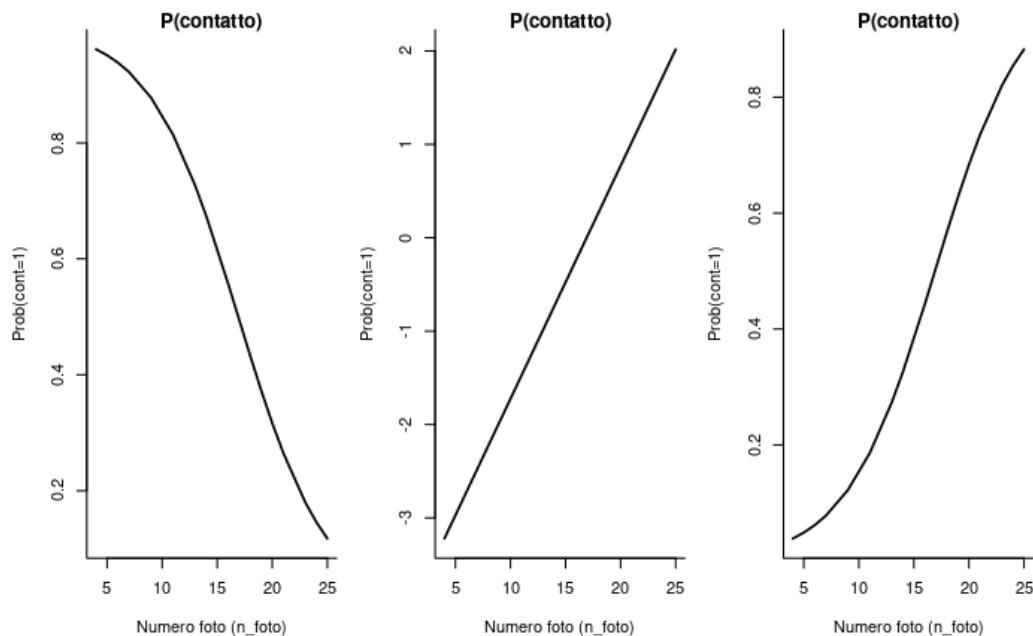
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.2205     1.6238  -2.60   0.0093 **
n_foto         0.2496     0.0903   2.76   0.0057 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 47.016  on 33  degrees of freedom
Residual deviance: 34.135  on 32  degrees of freedom
AIC: 38.14

Number of Fisher Scoring iterations: 5
```

La relazione stimata tra **n_foto** e **cont** viene mostrata in uno dei grafici sottostanti:



Viene poi derivato un test che verifica se β_1 , il coefficiente del modello che descrive l'impatto del numero di foto sulla probabilità di un contatto, è uguale a 0.1:

$$H_0 : \beta_1 = 0.1 \quad VS \quad H_1 : \beta_1 \neq 0.1$$

Inoltre si ha:

```
confint.default(fit_foto,"n_foto")
```

```
      2.5 % 97.5 %  
n_foto 0.07262 0.4265
```

Infine, viene stimato un secondo modello in cui tutte le variabili presenti nel dataset sono incluse come variabili esplicative:

```
fit_all <- glm(cont ~ ., data = df,  
              family = binomial)
```

e si desidera poi verificare quale modello fornisce una migliore bontà di adattamento:

```
logLik(fit_foto)
```

```
'log Lik.' -17.07 (df=2)
```

```
logLik(fit_all)
```

```
'log Lik.' -12.73 (df=5)
```

Si indichi quale grafico mostra la vera relazione tra n_foto e la stima della probabilità di osservare un incidente ottenuta nel modello `fit_foto`. Si indichi poi il valore della statistica test per la verifica di ipotesi indicata ($H_0 : \beta_1 = 0.1$) e si indichi se l'ipotesi può o non può essere rifiutata. Si indichi se il valore dell'AIC per il modello `fit_all` risulta maggiore o minore di quello del modello `fit_foto`. Infine si indichi cosa è possibile riferire sulla relazione tra il tipo di casa nell'annuncio e la probabilità che vi sia un contatto da parte di un cliente.

- a. ☐ La stima ottenuto nel modello `fit_foto` è mostrata nel pannello A.
☐ La stima ottenuto nel modello `fit_foto` è mostrata nel pannello B.
☐ La stima ottenuto nel modello `fit_foto` è mostrata nel pannello C.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: La stima ottenuto nel modello `fit_foto` è mostrata nel pannello C.

b. Il valore della statistica test per la verifica di ipotesi indicata: .

✗

c. ☐ L'ipotesi può essere rifiutata.

☐ L'ipotesi non può essere rifiutata.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: L'ipotesi non può essere rifiutata.

d. ☐ $AIC(\text{fit_foto}) > AIC(\text{fit_all})$.

☐ $AIC(\text{fit_foto}) < AIC(\text{fit_all})$.

☐ $AIC(\text{fit_foto}) = AIC(\text{fit_all})$.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: $AIC(\text{fit_foto}) > AIC(\text{fit_all})$.

e. ☐ La probabilità di un contatto è più alta per gli appartamenti.

☐ La probabilità di un contatto è più alta per gli case bi-familiari.

☐ Non è possibile affermare quale tipo di immobile ha la probabilità più alta di contatto..

☐ La probabilità di un contatto è più alta per le case singole.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: Non è possibile affermare quale tipo di immobile ha la probabilità più alta di contatto..

La probabilità che avvenga un contatto deve per definizione avere un valore in (0,1) - questo esclude il pannello in cui la funzione $P(\text{contatto})$ ha valori al di fuori dell'intervallo (0,1). Il coefficiente che descrive l'effetto di n_foto su $P(\text{Contatto})$ è positivo: questo indica che al crescere di n_foto cresce la probabilità di un contatto. Il livello di significatività usato è $\alpha = 0.05$.

La statistica test per un test quale

$$H_0: \beta_1 = \tilde{\beta} \quad VS \quad H_1: \beta_1 \neq \tilde{\beta}$$

si calcola con

$$\frac{\hat{\beta} - \tilde{\beta}}{sd(\hat{\beta})} = \frac{0.2496 - 0.1}{0.0903}$$

Il valore assoluto della statistica test va poi confrontato con il 97.5% quantile di una normale $z_{0.975} = 1.96$. In alternativa si può verificare se il valore 0.1 è contenuto nell'intervallo di confidenza (95%): se il valore è nell'intervallo l'ipotesi non può essere rifiutata.

$$AIC(\mathcal{M}(p)): -2 * \log Lik(\mathcal{M}(p)) + \log(n) * p$$

quindi

$$AIC(\text{fit_foto}) = 38.1354$$

$$AIC(\text{fit_all}) = 35.455$$

[Si potrebbe anche "semplicemente" controllare se la differenza tra la verosimiglianza di fit_all è maggiore o minore della differenza di gradi di libertà tra i due modelli (3).]

Non abbiamo elementi per giudicare quale tipo di casa ha un più alto tasso di contatto. L'informazione sarebbe deducibile avendo i valori dei coefficienti legati alla variabile **Type**, ma questi non sono disponibili.

- a. Falso / Falso / Vero
- b. 1.66
- c. Falso / Vero
- d. Vero / Falso / Falso
- e. Falso / Falso / Vero / Falso

Domanda 4

Risposta non data

Punteggio max.: 5,00

L'addetto alla sicurezza sul lavoro di un gruppo industriale monitora il numero di incidenti avvenuti nelle 4 diverse fabbriche del gruppo. Per ogni fabbrica (variabile FAB) è disponibile il numero di incidenti settimanali (Y) e il numero totale di ordini evasi nella settimana in decine (TE).

summary(df)

y	fab	te
Min. : 0.00	Bari :15	Min. :20.4
1st Qu.: 4.00	Nola :15	1st Qu.:28.1
Median : 6.00	Rho :15	Median :36.2
Mean : 7.93	Torino:15	Mean :38.0
3rd Qu.: 9.25		3rd Qu.:47.9
Max. :24.00		Max. :59.5

Si desidera indagare se il numero di ordini evasi ha un effetto sul numero di incidenti e l'importanza di questo effetto nelle diverse fabbriche. Il primo modello preso in esame è `fit_add`, un GLM (Poisson) in cui il predittore lineare include l'effetto di (te) e (fab) in maniera additiva:

```
fit_add <- glm(y~te+fab, family = poisson)
summary(fit_add)$coef
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.19448	0.189508	1.0262	3.048e-01
te	0.04470	0.004065	10.9980	3.905e-28
fabNola	0.36333	0.131346	2.7662	5.671e-03
fabRho	-0.39672	0.155254	-2.5553	1.061e-02
fabTorino	0.04365	0.137603	0.3173	7.511e-01

fit_add\$df.residual

[1] 55

Usando questo modello si desidera capire quale fabbrica ha il più alto numero di incidenti a parità di ordini evasi.

Si desidera poi valutare se vi sia evidenza che l'effetto della variabile (te) sia diverso da fabbrica a fabbrica. Per valutare l'ipotesi viene stimato il modello `fit_mult` e viene poi applicato un LRT, il cui p-value è riportato:

```
fit_mult <- glm(y~te*fab, family = poisson)
anova(fit_add, fit_mult, test = "LRT")$`Pr(>Chi)`[2]
```

[1] 0.4261

L'addetto desidera infine stimare, usando il modello `fit_add`, il numero di incidenti in una settimana in cui $(te=40)$ nelle fabbriche di Bari e Torino:

fit_add\$coef

(Intercept)	te	fabNola	fabRho	fabTorino
0.19448	0.04470	0.36333	-0.39672	0.04365

predict(fit_add, data.frame(nd = "Bari", te = 40),type = "response")

[1] NA

predict(fit_add, data.frame(nd = "Torino", te = 40),type = "response")

[1] NA

Si indichi quale fabbrica ha il più alto numero di incidenti a parità di ordini evasi. Si indichi quale interpretazione dare al risultato del LRT. Si indichino i valori stimati del numero di incidenti nelle fabbriche di Bari e Torino (cioè i valori mancanti delle funzioni `predict`). Infine si indichi il numero di gradi di libertà residui del modello `fit_mult` (cioè l'output di `fit_mult$df.residual`).

Il livello di significatività usato per i test è $\alpha = 0.05$.

- a. ☐ A parità di ordini evasi, Bari ha in media il maggior numero di incidenti.
☐ A parità di ordini evasi, Nola ha in media il maggior numero di incidenti.
☐ A parità di ordini evasi, Rho ha in media il maggior numero di incidenti.
☐ A parità di ordini evasi, Torino ha in media il maggior numero di incidenti.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: A parità di ordini evasi, Nola ha in media il maggior numero di incidenti.

- b. ☐ Non si può affermare che ci sia evidenza che l'effetto di TE sul numero di incidenti sia diverso nelle fabbriche .
☐ Vi è evidenza che l'effetto di TE sul numero di incidenti sia diverso nelle fabbriche.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: Non si può affermare che ci sia evidenza che l'effetto di TE sul numero di incidenti sia diverso nelle fabbriche .

- c. Il valore stimato di incidenti a Bari usando il modello `fit_add` quando `te=40`.

✗

- d. Il valore stimato di incidenti a Torino usando il modello `fit_add` quando `te=40`.

✗

- e. Il numero di gradi di libertà residui per `fit_mult`.

✗

A parità di ordini evasi la fabbrica con il valore di intercetta più alto ha il maggior numero di incidenti. Il valore dell'intercetta per la fabbrica di Bari corrisponde al valore dell'intercetta del modello, mentre per le altre fabbriche è necessario sommare il valore del coefficiente per i singoli livelli visibile nell'output di `summary`. Valori negativi del coefficiente indicano fabbriche in cui a parità del valore di `te` ci sono meno incidenti rispetto a Bari. Valori positivi del coefficiente indicano fabbriche in cui a parità del valore di `te` ci sono più incidenti rispetto a Bari.

La differenza tra il modello `fit_add` e `fit_mult` è che in `fit_mult` l'effetto di `te` sul numero di incidenti può essere diverso per ogni fabbrica: un valore alto del LRT (cioè un valore basso del `pvalue`) indicherebbe una forte evidenza che l'effetto di `te` sia effettivamente diverso nelle varie fabbriche. Dato che il modello `fit_mult` aggiunge tre possibili diversi coefficienti che descrivono l'effetto di `te` sugli incidenti nelle fabbriche che non sono Bari il numero di gradi di libertà residui del modello è: `fit_add$df.residual - 3`, 52.

Il modello usa il coefficiente canonico (cioè il default), che per la distribuzione di Poisson è il logaritmo. Si ha quindi:

- il valore stimato a Bari: $\exp\{0.1945 + 1.7882\}$
- il valore stimato a Torino: $\exp\{0.1945 + (0.0437) + 1.7882\}$

- a. Falso / Vero / Falso / Falso
b. Vero / Falso
c. 7.26

d. 7.59

e. 52.00

Domanda 5

Risposta non data

Punteggio max.: 3,00

L'addetta alla sicurezza sul lavoro di una fabbrica tiene un record in cui indica se in una giornata sono registrati incidenti sul lavoro (una variabile dicotomica Y , in cui $Y=1$ indica che vi è stato un incidente). L'addetta desidera indagare se vi sono delle variabili esterne che influenzano la probabilità di registrare un incidente sul lavoro. Raccoglie quindi informazioni su potenziali fattori esterni quali il numero totale di ordini evasi nella giornata (TE) e alcune variabili climatiche (X_1, X_2, X_3).

Il primo modello usato per indagare l'effetto del numero di ordini evasi sulla probabilità che avvenga almeno un incidente nell'impianto è un GLM (Binomiale) in cui solo la variabile TE viene inserita nel predittore lineare:

```
fit1 <- glm(y~te, data = df, family = binomial)
summary(fit1)
```

Call:

```
glm(formula = y ~ te, family = binomial, data = df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.117	-0.341	-0.127	0.447	1.893

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	14.960	3.727	4.01	6.0e-05 ***
te	-0.512	0.127	-4.03	5.5e-05 ***

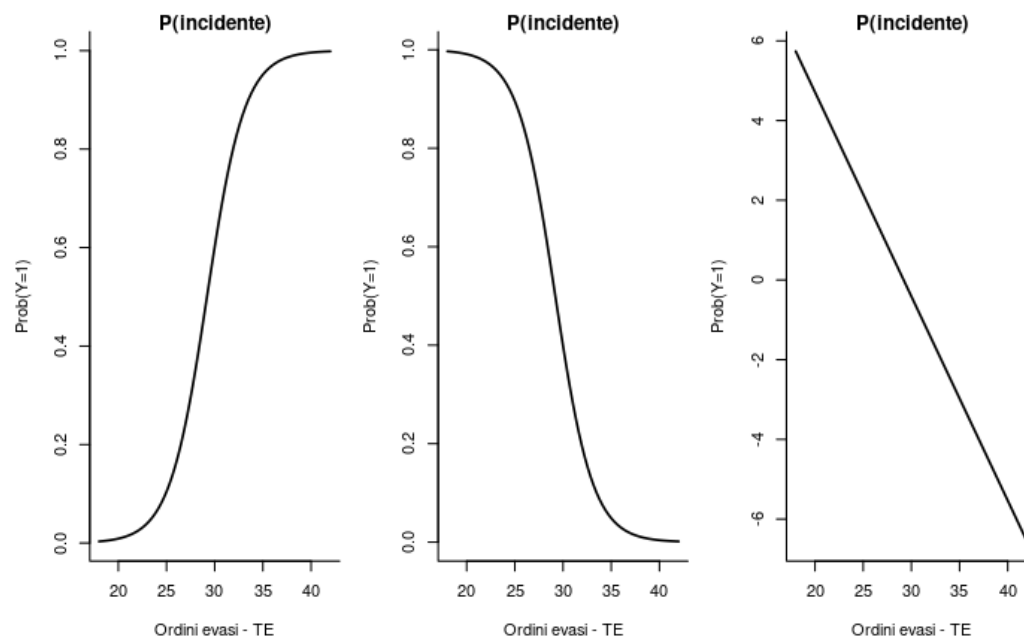
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 82.577 on 59 degrees of freedom
 Residual deviance: 40.230 on 58 degrees of freedom
 AIC: 44.23

Number of Fisher Scoring iterations: 6

La relazione stimata TE e Y dal modello viene mostrata in uno dei grafici sottostanti:



Viene poi derivato l'intervallo di confidenza al 95% per il parametro β_0 (l'intercetta del modello):

```
confint.default(fit_te, "(Intercept)")
```

```
      2.5 % 97.5 %  
(Intercept) 7.655 22.27
```

e si desidera poi verificare se un modello in cui vengono inserite anche le variabili climatiche fornisce una migliore bontà di adattamento:

```
fit_all <- glm(y~te+x1+x2+x3,family = binomial)  
deviance(fit_te)
```

```
[1] 40.23
```

```
deviance(fit_all)
```

```
[1] NA
```

Si indichi quale grafico mostra la vera relazione tra te e la stima della probabilità di osservare un incidente ottenuta nel modello `fit_te`. Si indichi poi quale distribuzione viene usata per derivare l'intervallo di confidenza di β_0 . Si indichi infine se il valore della devianza del modello `fit_all` è maggiore o minore del valore della devianza del modello `fit_te`.

- a. ☐ La stima ottenuto nel modello `fit_te` è mostrata nel pannello A.
☐ La stima ottenuto nel modello `fit_te` è mostrata nel pannello B.
☐ La stima ottenuto nel modello `fit_te` è mostrata nel pannello C.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: La stima ottenuto nel modello `fit_te` è mostrata nel pannello B.

- b. ☐ Per costruire intervalli di confidenza per β_0 si usa una distribuzione Normale.
☐ Per costruire intervalli di confidenza per β_0 si usa una distribuzione T-Student.
☐ Per costruire intervalli di confidenza per β_0 si usa una distribuzione F.
☐ Per costruire intervalli di confidenza per β_0 si usa una distribuzione Chi-quadro.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: Per costruire intervalli di confidenza per β_0 si usa una distribuzione Normale.

- c. ☐ `deviance(fit_all) < deviance(fit_te)`.
☐ `deviance(fit_all) > deviance(fit_te)`.

Punteggio ottenuto 0,00 su 1,00

La risposta corretta è: `deviance(fit_all) < deviance(fit_te)`.

Dato che il coefficiente che descrive la relazione tra `te` e la probabilità di osservare un incidente è negativo si evince che al crescere di `te` diminuisce la probabilità di osservare un incidente. Inoltre il grafico deve mostrare nella asse delle ordinate valori tra (0,1) e la tipica forma della curva logistica. La stima ottenuto nel modello `fit_te` è mostrata nel pannello B.

L'inferenza per i parametri nei GLM si basa sul fatto che le stime dei coefficienti di regressione sono stime di massima verosimiglianza e sono di conseguenza approssimativamente normalmente distribuite (per $n \rightarrow \infty$).

Più si aggiungono parametri al modello (cioè più si aumenta la complessità del modello) più il modello catturerà la variabilità dei dati (sebbene con il rischio che il miglioramento in termini di bontà di adattamento del modello sia minimo rispetto al costo di includere ulteriori parametri). Per questo motivo si ha che: $\text{deviance}(\text{fit_all}) < \text{deviance}(\text{fit_te})$.

- a. Falso / Vero / Falso
- b. Vero / Falso / Falso / Falso
- c. Vero / Falso

◀ Quiz Modelli Regressione Multipla

Vai a...

Esercizi Modelli Lineari ►